# **15.095 Project Proposal:** Predict Demand to Prescribe Ingredient Orders

*November 07th, 2025*

*Karl Chaker (kchaker8@mit.edu), Matthieu Hakim (mhakim@mit.edu)*

## Problem Summary

This project aims to develop a prescriptive analytics framework to optimize ingredient purchasing decisions for a pizza restaurant. Using historical transactional sales and a constructed recipe matrix, we will forecast demand for each menu item by day and daypart (Lunch, Afternoon, Dinner) and prescribe ingredient order quantities (kg/case) that balance profitability, service levels, and operational constraints. By integrating predictive modeling with optimization, we will demonstrate how data-driven decision-making can improve inventory management and margins under operational uncertainty.

*The same framework extends to other allocation settings, e.g. Central kitchens, food distributors, and food banks/shelters.*

## Dataset Description (https://www.kaggle.com/datasets/neethimohan/maven-pizza-challenge-dataset)

Primary dataset: Kaggle, Maven Pizza Challenge (orders, order_details, pizzas, pizza_types), which provides ticket timestamps, menu items, sizes, prices, categories, and ingredients. We will augment this with: (i) an ingredient cost catalog (per kg or per case) with pack sizes, (ii) ingredient usage factors per pizza type & size (e.g., Large Pepperoni consumes 0.20 case-equivalents of pepperoni, with size multipliers S/M/L/XL), (iii) external features including holiday indicators, simple local weather, and an oil-price proxy for inflation, and (iv) time-of-day bucketing that maps timestamps to Lunch/Afternoon/Dinner.

## Methodology

Predictive: Build a supervised learning model to forecast demand for each menu item per day and daypart. Features include day-of-week, month/season, daypart, holiday flags, oil price and sparse lag features (e.g., prior-week same-day, short moving averages). Models will emphasize tree-based methods (CART/Random Forest/XGBoost/ORT) with shallow interpretable benchmarks when appropriate.

Prescriptive: Formulate a multi-product newsvendor over ingredients. Decision variables are ingredient order quantities (kg/cases). Objective: maximize profit (sales margin) minus penalties for unmet demand (stockouts) and spoilage/waste. Constraints ensure: (a) recipe feasibility (purchased ingredients cover forecasted pizzas via the recipe matrix), (b) storage capacity, and (c) pack sizes/minimum order quantities; option to include shelf-life bounds for perishables.

Weighted-cost prescriptions: Compare standard point-prediction with a data-weighted approach that minimizes historical cost over 'similar-day' neighborhoods. We will experiment with K-Means clustering and Optimal Regression Trees (as a partitioner) to compute cluster-weighted costs and corresponding prescriptions.

## Challenges and Potential Solutions

A key challenge will be the estimation of ingredient costs and quantities, which are not directly available in the dataset. We plan to approximate these values using reasonable market-based assumptions or publicly available pricing data. Another challenge involves model calibration and the computational complexity of solving large-scale optimization problems. To address this, we will perform feature selection and, if necessary, aggregate similar menu items to simplify the model while preserving interpretability. Finally, we will need to address forecast uncertainty and distribution shifts (e.g., holidays/weather) via neighborhood-weighted prescriptions and scenario/sensitivity analysis. Through this project, we aim to combine predictive modeling and optimization techniques from class to create a practical and interpretable decision-support tool for restaurant inventory management.